# "Internal Linear Combination" method for the separation of CMB from Galactic foregrounds in the harmonic domain

R. Vio[1] and P. Andreani[2]

[1] Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy
   e-mail: robertovio@tin.it,
[2] ESO, Karl Schwarzschild strasse 2, 85748 Garching, Germany
   INAF-Osservatorio Astronomico di Trieste, via Tiepolo 11, 34143 Trieste, Italy
   e-mail: pandrean@eso.org

**ABSTRACT**

Foreground contamination is the fundamental hindrance to the cosmic microwave background (CMB) signals and its separation from it represents a fundamental question in Cosmology. One of the most popular algorithm used to disentangle foregrounds from the CMB signals is the "*internal linear combination*" method (ILC). In its original version, this technique is applied directly to the observed maps. In recent literature, however, it is suggested that in the harmonic (Fourier) domain it is possible to obtain better results since a separation can be attempted where the various Fourier frequencies are given different weights. This is seen as a useful characteristic in the case of noisy data. Here, we argue that the benefits of using such an approach are overestimated. Better results can be obtained if a classic procedure is adopted where data are filtered before the separation is carried out.

**Key words.** Methods: data analysis – Methods: statistical – Cosmology: cosmic microwave background

## 1. Introduction

The experimental progresses in the detection of cosmological emissions require a parallel development of data analysis techniques in order to extract the maximum physical information from data. In particular, different emission mechanisms are characterized by markedly distinct underlying physical processes. Data analysis often requires the component separation in order to study the individual characteristics. To achieve such a goal, a link between the branch of signal processing science which characterizes and separates different signals and astrophysics is yet well established, and in many cases, modern signal processing techniques have been imported and applied in an astrophysical context.

In this work, we consider one of the most widely used approaches for the separation of different emissions, say the *internal linear combination* method (ILC), which was adopted, for instance, in the reduction of the data from the Wilkinson Microwave Anisotropy Probe (WMAP) satellite for CMB observations (Bennett et al. 2003). Among the separation techniques, ILC requires the smallest number of *a priori* assumptions. Here, the available data are assumed in the form of $N_o$ maps, taken at different frequencies, containing $N_p$ pixels each. More precisely, if $S^{(i)}(p)$ provides the value of the $p$th pixel for a map obtained at channel "$i$"[1], our starting model is:

$$S^{(i)}(p) = \mathfrak{S}_c^{(i)}(p) + S_f^{(i)}(p) + \mathcal{N}^{(i)}(p) \qquad (1)$$

---

[1] In the present work, $p$ indexes pixels in the classic spatial domain. However, the same formalism applies if other domains are considered, for example, the Fourier one.

where $\mathfrak{S}_c^{(i)}(p)$, $S_f^{(i)}(p)$ and $\mathcal{N}^{(i)}(p)$ are the contributions due to the CMB, the diffuse Galactic foreground and the experimental noise, respectively. Although not necessary for later arguments, it is assumed that all of these contributions are representable by means of stationary random fields. Moreover, without loss of generality, for ease of notation the random fields are supposed as the realization of zero-mean spatial processes. In the present work the contribution of non-diffuse components (e.g., due to SZ cluster, point-sources, . . . ) are not considered and they are assumed to have been removed through other methodologies.

The idea behind ILC is simple. The main assumption is that model (1) can be written as

$$S^{(i)}(p) = \mathfrak{S}_c(p) + S_f^{(i)}(p) + \mathcal{N}^{(i)}(p), \qquad (2)$$

i.e. the template of the CMB component is independent of the observing channel. A way to exploit this fact is to average $N_o$ images $\{S^{(i)}(p)\}_{i=1}^{N_o}$ giving a specific weight $w_i$ to each of them so as to minimize the impact of the foreground and noise (Bennett et al. 2003). This means to look for a solution of type

$$\widehat{\mathfrak{S}}_c(p) = \sum_{i=1}^{N_o} w_i S^{(i)}(p). \qquad (3)$$

If the constraint $\sum_{i=1}^{N_o} w_i = 1$ is imposed, Eq. (3) becomes

$$\widehat{\mathfrak{S}}_c(p) = \mathfrak{S}_c(p) + \sum_{i=1}^{N_o} w_i[S_f^{(i)}(p) + \mathcal{N}^{(i)}(p)]. \qquad (4)$$

Now, from this equation it is clear that, for a given pixel "$p$", the only variable terms are in the sum. Hence, under

the assumption of independence of $\mathfrak{S}_c(p)$ from $S_f^{(i)}(p)$ and $\mathcal{N}^{(i)}(p)$, the weights $\{w_i\}$ have to minimize the variance of $\widehat{\mathfrak{S}}_c(p)$, i.e.

$$\{w_i\} = \underset{\{w_i\}}{\arg\min}$$

$$\text{VAR}\left[\mathfrak{S}_c(p)\right] + \text{VAR}\left[\sum_{i=1}^{N_o} w_i(S_f^{(i)}(p) + \mathcal{N}^{(i)}(p))\right], \quad (5)$$

where $\text{VAR}[s(p)]$ is the *expected variance* of $s(p)$. If $\boldsymbol{S}^{(i)}$ denotes a **row vector** such as $\boldsymbol{S}^{(i)} = [S^{(i)}(1), S^{(i)}(2), \ldots, S^{(i)}(N_p)]$ and the $N_o \times N_p$ matrix $\boldsymbol{S}$ is defined as

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}^{(1)} \\ \boldsymbol{S}^{(2)} \\ \vdots \\ \boldsymbol{S}^{(N_o)} \end{pmatrix}, \quad (6)$$

then Eq. (2) becomes

$$\boldsymbol{S} = \mathfrak{S}_c + \boldsymbol{S}_f + \mathcal{N}. \quad (7)$$

In this case, the weights are given by (Eriksen et al. 2004)

$$\boldsymbol{w} = \frac{\boldsymbol{C}_S^{-1}\mathbf{1}}{\mathbf{1}^T \boldsymbol{C}_S^{-1}\mathbf{1}}, \quad (8)$$

where $\boldsymbol{C}_S$ is the $N_o \times N_o$ cross-covariance matrix of the random processes that generate $\boldsymbol{S}$, i.e.

$$\boldsymbol{C}_S = \text{E}[\boldsymbol{S}\boldsymbol{S}^T], \quad (9)$$

and $\mathbf{1} = (1, 1, \ldots, 1)^T$ is a column vector of all ones. Here, E[.] denotes the *expectation operator*. Hence, the ILC estimator takes the form

$$\widehat{\mathfrak{S}}_c = \boldsymbol{w}^T \boldsymbol{S}, \quad (10)$$

$$= \alpha \mathbf{1}^T \boldsymbol{C}_S^{-1} \boldsymbol{S}, \quad (11)$$

with $\mathbf{1}^T \boldsymbol{w} = 1$ and the scalar quantity $\alpha$ given by

$$\alpha = [\mathbf{1}^T \boldsymbol{C}_S^{-1}\mathbf{1}]^{-1}. \quad (12)$$

In practical applications, matrix $\boldsymbol{C}_S$ is unknown and has to be estimated from the data. Typically, this is done by means of the estimator

$$\widehat{\boldsymbol{C}}_S = \frac{1}{N_p} \boldsymbol{S}\boldsymbol{S}^T. \quad (13)$$

A common assumption in CMB observations is that $S_f^{(i)}$ is given by the linear mixture of the contribution of $N_c$ physical processes $\{\mathfrak{S}_j\}_{j=1}^{N_c}$ (e.g., free-free, dust re-radiation, . . . )

$$\boldsymbol{S}_f^{(i)} = \sum_{j=1}^{N_c} a_{ij} \mathfrak{S}_j, \quad (14)$$

with $a_{ij}$ constant coefficients. In practice, it is assumed that for the $j$th physical process a template $\mathfrak{S}_j$ exists independent of the specific channel " $i$ ". Although rather strong, it is not unrealistic to assume that this condition is satisfied when small enough patches of the sky are considered. Inserting Eq. (14) into Eq. (7) one obtains

$$\boldsymbol{S} = \boldsymbol{A}\mathfrak{S} + \mathcal{N} \quad (15)$$

with

$$\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_c \\ \mathfrak{S}_1 \\ \mathfrak{S}_2 \\ \vdots \\ \mathfrak{S}_{Nc} \end{pmatrix}, \quad (16)$$

and

$$\boldsymbol{A} = \begin{pmatrix} 1 & a_{11} & a_{12} & \ldots & a_{1N_c} \\ 1 & a_{21} & a_{22} & \ldots & a_{2N_c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{N_o1} & a_{N_o2} & \ldots & a_{N_oN_c} \end{pmatrix}. \quad (17)$$

Here, matrix $\boldsymbol{A}$ is assumed to be of full rank.

## 2. ILC in the spatial domain

In a recent paper, Vio & Andreani (2008) have stressed various problems concerning ILC that in literature have been underevaluated if not undetected. For example, in Eq. (7) the term $\boldsymbol{S}_f + \mathcal{N}$ is often considered as a single noise component (e.g., see Eriksen et al. 2004; Hinshaw et al. 2007; Kim et al. 2007, 2008). In this way the problem is apparently simplified since it is reduced to the separation of two components only. No *a priori* information on this "global" noise is required. However, this approach can lead to wrong conclusions. For example, since all the components in the mixtures $\boldsymbol{S}$ are assumed to be zero-mean, from Eq. (4) one could conclude that

$$\text{E}[\widehat{\mathfrak{S}}_c | \mathfrak{S}_c] = \mathfrak{S}_c + \boldsymbol{w}^T \text{E}[\boldsymbol{S}_f + \mathcal{N}] = \mathfrak{S}_c, \quad (18)$$

i.e. the ILC estimator is unbiased [2]. This is not correct: the claim that $\widehat{\mathfrak{S}}_c$ is unbiased requires one to prove that

$$\text{E}[\widehat{\mathfrak{S}}_c | \mathfrak{S}_c, \boldsymbol{S}_f] = \mathfrak{S}_c + \boldsymbol{w}^T \boldsymbol{S}_f + \boldsymbol{w}^T \text{E}[\mathcal{N}] = \mathfrak{S}_c. \quad (19)$$

The reason is that $\boldsymbol{S}_f$ is a fixed realization of a random process. There is no way to obtain another one. Even if observed many times (under the same experimental conditions) the foreground components (for instance the Galaxy) will always appear the same. Only the noise component $\mathcal{N}$ will change. This has important consequences. In fact, in Vio & Andreani (2008) it is proved that, even under model (15), ILC can provide satisfactory results only under rather restrictive conditions:

1. The number of observing frequencies $N_o$ has to be larger than the number of components $N_c$;

2. $\mathfrak{S}_c$ has to be uncorrelated with $\boldsymbol{S}_f$;

3. Data have to be noise-free, i.e. $\mathcal{N} = \mathbf{0}$.

The violation of any of these points has as consequence the introduction of a bias in the ILC solution that can be severe. In literature it appears that only the second point has been well realized. (e.g. see Delabrouille & Cardoso 2007). The explanation of the bias when $N_o \leq N_c$ is technical and we remand to Vio & Andreani (2008). When noise is present, under the condition of $\mathcal{N}$ uncorrelated with $\boldsymbol{S}$, from model (15) it is

$$\boldsymbol{C}_S = \boldsymbol{A}\boldsymbol{C}_{\mathfrak{S}}\boldsymbol{A}^T + \boldsymbol{\Omega}_{\mathcal{N}}, \quad (20)$$

---

[2] The expression $\text{E}[a|b]$ indicates *conditional expectation* of $a$ given $b$.

with

$$\boldsymbol{\Omega}_{\boldsymbol{\mathcal{N}}} = \mathrm{E}[\boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{N}}^T]. \tag{21}$$

Hence, a bias derives from the fact that $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{N}}}$ is a matrix with strictly positive entries and then $\mathrm{E}[\widehat{\boldsymbol{C}}_{\boldsymbol{S}}] \neq \boldsymbol{C}_{\boldsymbol{S}}$.

In absence of "*a priori information*", the problems connected to the first two points above have no solution. The only possibility is to plan the experiments in such a way to avoid them (e.g. observations at high Galactic latitudes, a number of observing frequencies sufficiently large ... ). On the other hand, noise is an unavoidable question that, however, can be handled with hope of satisfactory results. In this respect, some authors (e.g. see Tegmark et al. 2003; Kim et al. 2007, 2008) suggest that an effective way is to use ILC in the harmonic (Fourier) domain. In practice, this means to apply ILC to $\widetilde{\boldsymbol{S}}$ that is the matrix whose $i$th row contains the two-dimensional Fourier transform of $\boldsymbol{S}^{(i)}$. Following Kim et al. (2007), we will indicate this version of ILC as "*harmonic internal linear combination*" (HILC).

## 3. ILC in the Fourier domain (HILC)

The starting consideration is that all arguments presented in the previous sections hold independently of the fact that one is working in the ordinary spatial domain or in the Fourier domain. Indeed, it is not difficult to see that the weights $\boldsymbol{w}$ as given by Eq. (8) can be equivalently computed by means of

$$\boldsymbol{w} = \frac{\boldsymbol{C}_{\widetilde{\boldsymbol{S}}}^{-1}\boldsymbol{1}}{\boldsymbol{1}^T\boldsymbol{C}_{\widetilde{\boldsymbol{S}}}^{-1}\boldsymbol{1}}. \tag{22}$$

In this respect, it is sufficient to take into account that

$$\widetilde{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{\mathcal{F}}, \tag{23}$$

with $\boldsymbol{\mathcal{F}} = \boldsymbol{F}_{N_c} \otimes \boldsymbol{F}_{N_o}$, "$\otimes$" the Kronecker product, $\boldsymbol{F}_N$ the $N \times N$ the Fourier matrix that is a complex, symmetric and unitary matrix whose elements are given by

$$(F_N)_{kl} = \frac{1}{\sqrt{N}}\mathrm{e}^{-2\pi\iota(k-1)(l-1)/N}, \tag{24}$$

$\iota = \sqrt{-1}$. Hence

$$\boldsymbol{C}_{\widetilde{\boldsymbol{S}}} = \mathrm{E}[\widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{S}}^H] = \mathrm{E}[\boldsymbol{S}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{S}^T] = \boldsymbol{C}_{\boldsymbol{S}}, \tag{25}$$

since $\boldsymbol{F}\boldsymbol{F}^H = \boldsymbol{F}^H\boldsymbol{F} = \boldsymbol{I}$ where $\boldsymbol{F}_N^H$ is the complex conjugate transpose of $\boldsymbol{F}_N$.

It is well known that measurement noise $\{\boldsymbol{\mathcal{N}}^{(i)}\}$, typically the realization of wide-band stochastic processes, tends to uniformly spread in the Fourier domain. On the other hand, the contribution of signals $\{\boldsymbol{S}^{(i)}\}$ is typically concentrated in correspondence to the lowest Fourier frequencies. For example, this is visible in Fig. 1 where the power spectra of three simulated mixtures are shown (see Fig. 3). Here, non-astronomical signals have been used, but the same holds also for astronomical ones. This implies that, for the indices "$p$" corresponding the lowest frequencies, the contribution of noise to $\widehat{S}^{(i)}(p)$ is quite small. Hence, the basic idea here is to partition the frequency domain in a number $N_s$ of subsets and to apply separately the ILC filter to each of them. At the end of this procedure an estimate $\widehat{\boldsymbol{\mathfrak{S}}}_c \equiv \bigcup_{k=1}^{N_s} \widetilde{\boldsymbol{\mathfrak{S}}}_c(p_k)$ is obtained where

$p_k = \{p|p \in k\text{-}th\ frequency\ subset\}$. Finally, $\widehat{\boldsymbol{\mathfrak{S}}}_c$ can be recovered by the Fourier inversion of $\widetilde{\boldsymbol{\mathfrak{S}}}_c$.

The effectiveness of such an approach seems supported by the simple example presented in Figs. 2-4. In particular, Fig. 2 shows the original images $\boldsymbol{\mathfrak{S}}_c$, $\boldsymbol{\mathfrak{S}}_1$, $\boldsymbol{\mathfrak{S}}_2$ as well the mixing matrix $\boldsymbol{A}$ that, through model (15), are used to create the observed images $\{\boldsymbol{S}^{(i)}\}_{i=1}^3$ shown in Fig. 3. As reference for later results, the bottom-right panel in the same figure shows the estimate $\widehat{\boldsymbol{\mathfrak{S}}}_c$ obtained when ILC is applied to these images. For this case, the resulting *root mean square* (rms) of the residuals is about 0.05. The top panels in Fig. 4 show what happens when the images are added a white-noise with a *signal-to-noise* (SNR) [3] set to 5. Here, both ILC and HILC are used. In the case of HILC, the frequency domain is partitioned in two regions, as shown in the bottom-left panel of Fig. 1, one containing the low frequencies and the other containing the high frequencies. The weights $\boldsymbol{w}$ are calculated independently for each of them. It is evident that HILC outperforms ILC. This conclusion is confirmed by the *rms* of the its residuals that is 0.13 and 0.26, respectively. The same indication is provided by the bottom panels that show the result obtained when the weights computed for the noisy images are applied to the noise-free ones. This operation gives an idea of the bias introduced in the solution by the noise. In this case, the *rms* of the residuals become 0.08 and 0.18, respectively.

## 4. Discussion and conclusions

Reader should not be confused by the fact that at first sight HILC appears an approach potentially superior to ILC. It is indeed true that HILC performs a more effective separation of the signal from the noise. However, if the noise affects most the high frequency part of the signal, why then do not simply filter out this one? Indeed, the left panel of Fig. 5 shows the estimate $\widehat{\boldsymbol{\mathfrak{S}}}_c$ obtained when, before the application of ILC, the images $\{\boldsymbol{S}^{(i)}\}$ are filtered with the ideal Fourier circular low-pass filter [4] that has the structure identical to that shown in Fig. 1. As before, the right panel of this figure shows the solution obtained when the resulting weights are applied to the noise-free version of the data. The *rms* of the residuals are 0.12 and 0.08, respectively. The comparison with the values of 0.13 and 0.08, previously determined for HILC, indicates an improvement of the quality of the result. The indication that comes out from this simple experiment is that filtering noise is a more effective operation than its separation from the signal of interest. It is true that the version of HILC here used is rather rough and more sophisticated ones are available (Kim et al. 2007, 2008). However, the same holds also for the filtering operation that has been coupled with ILC. At this point a question arises: which is the benefit in using a non-standard approach as HILC with respect to a classic approach where filtering is coupled with ILC? This in not an academic question. Indeed, the use of non-standard tools implies the renunciation of a huge body of experience gained in years of application in practical prob-

---

[3] Here, the quantity SNR is defined as the ratio between the standard deviation of the signal with the standard deviation of the noise.

[4] In the Fourier domain, an ideal filter $\widetilde{\boldsymbol{H}}$ has the form $\widetilde{H}(p) = 1$ if $p$ is a frequency of interest, 0 otherwise.
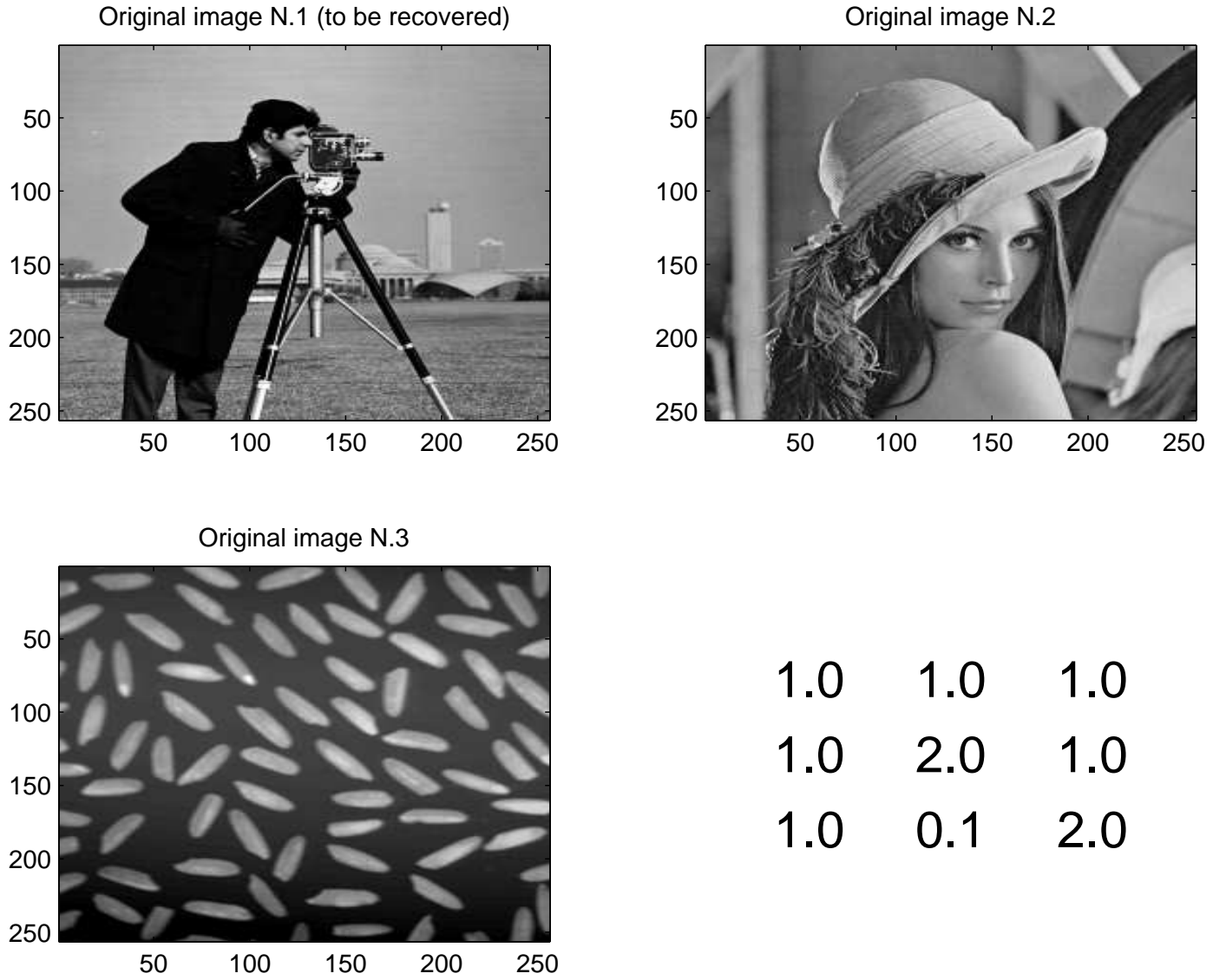
lems also in disciplines different from Astronomy. In other words, the choice of non-standard tools is indicated only in situations of real and sensible improvements of the results. New techniques that do not fulfill this requirement should be introduced with care: they prevent the comparison with the results obtained in other works and may lead people to use not well tested methodologies ending up in not reliable results.

## References

Bennett, C.L., et al. 2003, ApJS, 148, 97

Delabrouille, J., & Cardoso, J.F. 2007, arXiv:astro-ph/0702198v2

Eriksen, H.K., Banday, A.J., Górski, K.M., & Lilje, P.B. 2004, ApJ, 612, 633

Hinshaw, G., et al. 2007, ApJS, 170, 288

Kim, j., Naselsky, P., & Christensen, P.R. 2007, Phys. Rev. D, 375, 625

Kim, j., Naselsky, P., & Christensen, P.R. 2008, arXiv:0810.4008

Tegmark, M., de Oliveira-Costa, A., & Hamilton, A.J. 2003, Phys. Rev. D., 68, 123523

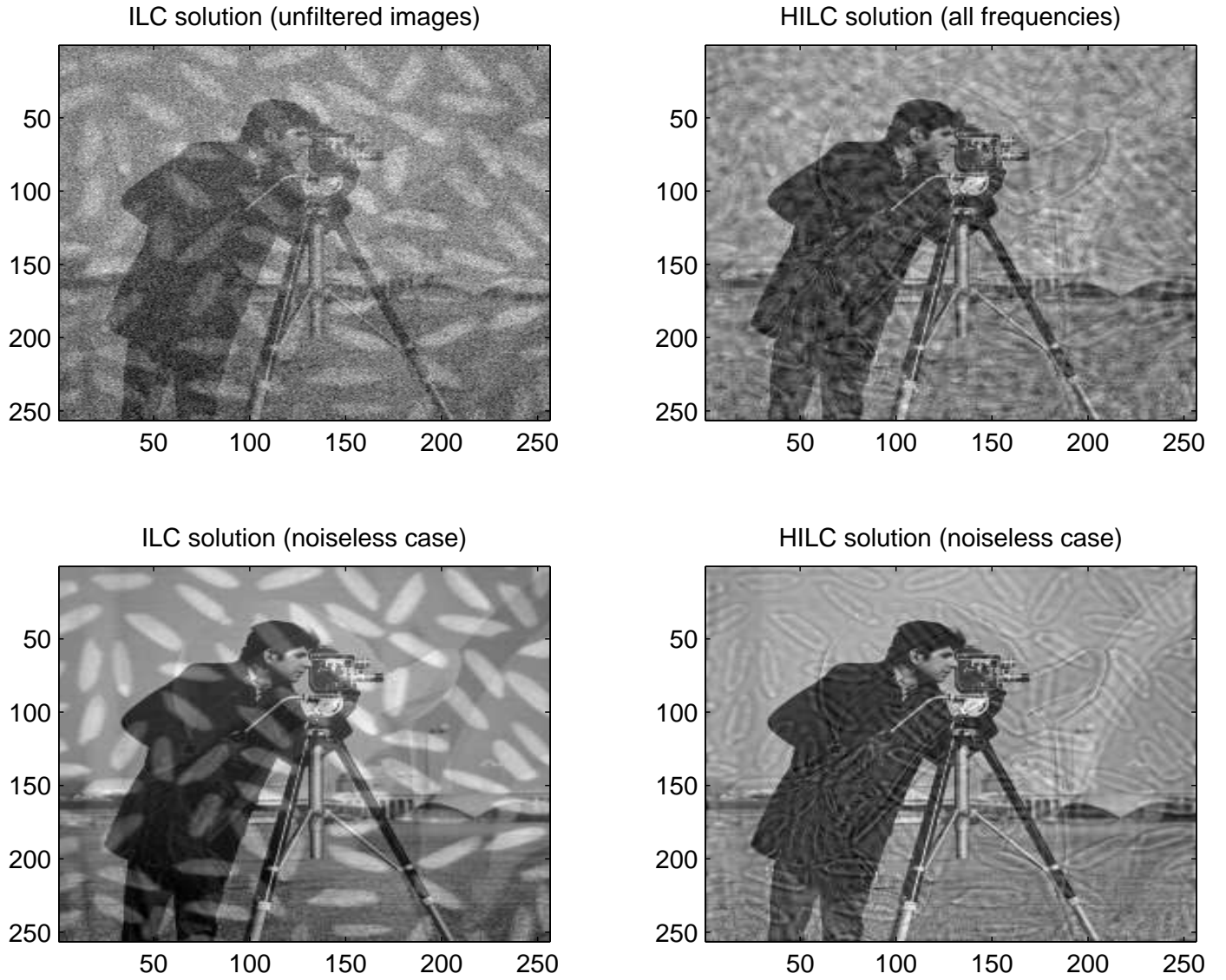Vio, R., & Andreani, P. 2008, A&A 487, 775
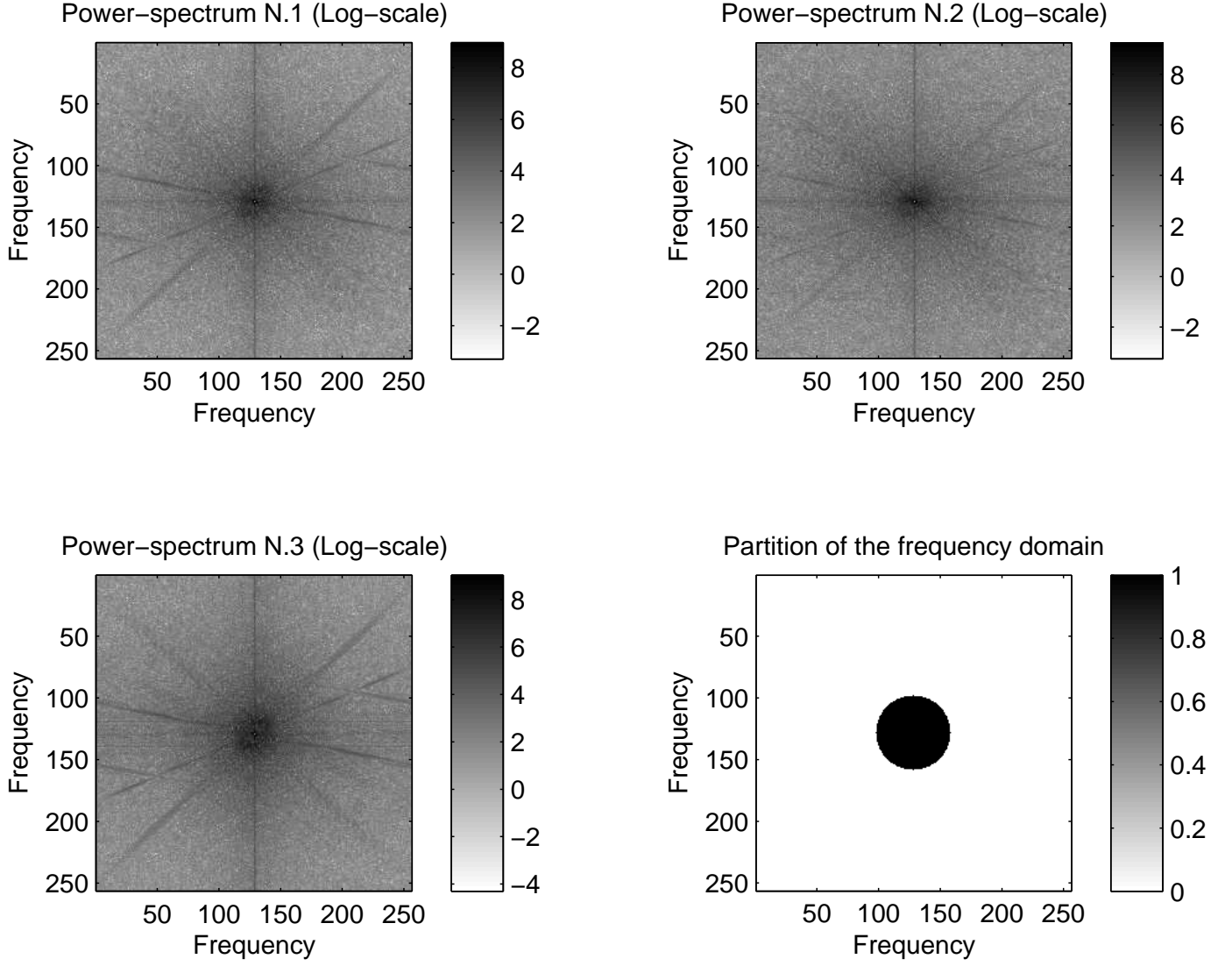
**Fig. 1.** Top panels and bottom-left panel: power spectra (logarithmic scale) of the images $\boldsymbol{S}^{(i)}$ shown in Fig. 3. The Figure shows that most of the power is concentrated at low frequency. Bottom-right panel: partition of the discrete Fourier domain that is used by HILC as discussed in Sect. 3.

Original image N.1 (to be recovered)

Original image N.2

Original image N.3

$$\begin{matrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 2.0 & 1.0 \\ 1.0 & 0.1 & 2.0 \end{matrix}$$

**Fig. 2.** Top panels and bottom-left panel: original images $\mathfrak{S}_c$, $\mathfrak{S}_1$ and $\mathfrak{S}_2$ – see Eq. (16) – used in the experiment described in the text. The top-left panel shows the image to recover. Through Eq. (15) these images are used to produce the mixtures $\boldsymbol{S}^{(i)}$ shown in Fig. 3. The bottom-right panel provides the mixing matrix $\boldsymbol{A}$ – see Eq. (17).
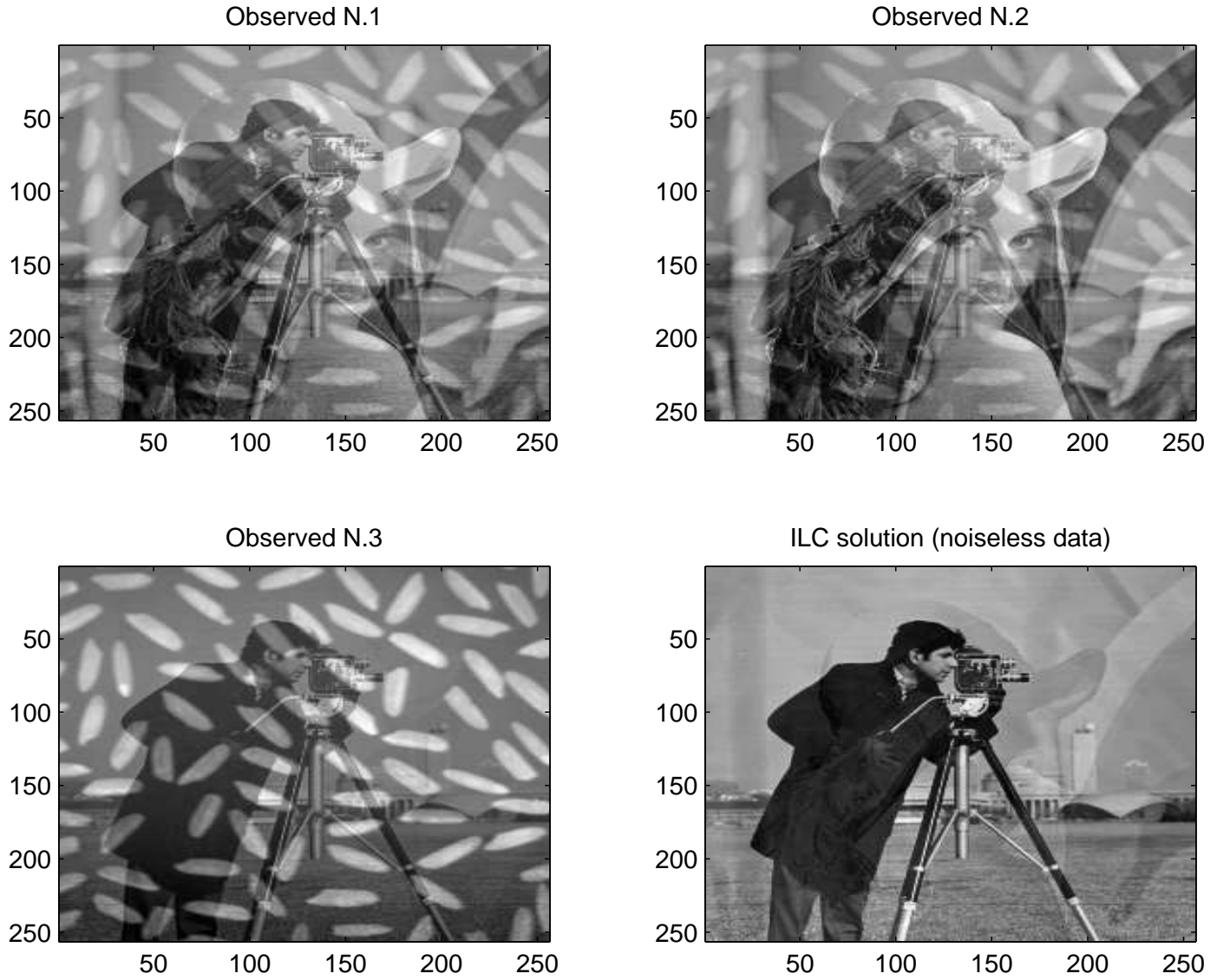
**Fig. 3.** Top panels and bottom-left panel: observed mixtures $\boldsymbol{S}^{(1)}$, $\boldsymbol{S}^{(2)}$ and $\boldsymbol{S}^{(3)}$ – see Eq. (15) – used in the experiment described in the text. The data shown here are noiseless. The right-bottom panel shows the solution obtained when ILC is applied to these images.

ILC solution (unfiltered images)

HILC solution (all frequencies)

ILC solution (noiseless case)

HILC solution (noiseless case)



**Fig. 4.** Top panels: solutions provided by ILC (left panel) and HILC (right panel) - see Sec. 3 when a white-noise, with a signal-to-noise ratio set to 5, is added to the images shown in Fig. 3. Bottom panels: Results obtained when the weights used to compute the solutions shown in the top panels are applied to the original (i.e. noise-free) images in Fig. 3). This operation gives an idea of the bias introduced in the solution by the noise.

ILC solution (filtered images)                           ILC solution (noiseless case)
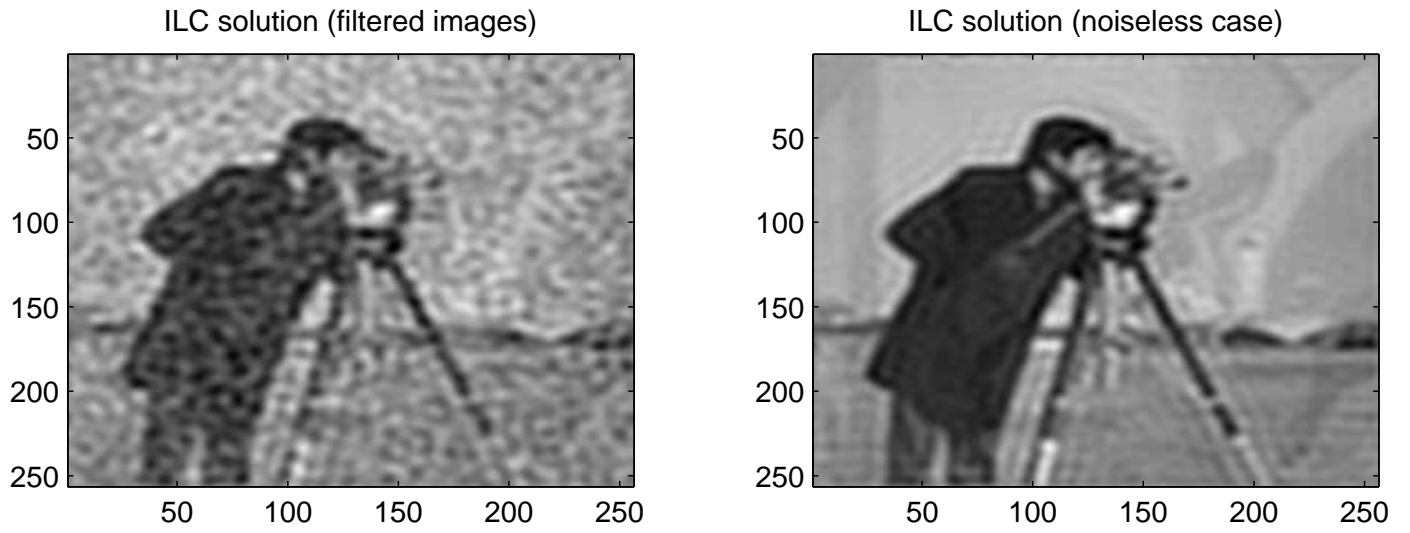


**Fig. 5.** As in the left panels of Fig. 4 with the difference that before using ILC, the images have been filtered with the ideal circular low-pass filter that has the same structure as in the bottom-right panel of Fig. 1.